

Psychometric properties of the Slovenian adaptation of the Revised Generic Occupational Stress Index Questionnaire (RG-OSI)

Nataša Sedlar^{1*}, Gregor Sočan² and Lilijana Šprah³

¹National Institute of Public Health, Ljubljana, Slovenia

²Department of Psychology, Faculty of Arts, University of Ljubljana, Slovenia

³Sociomedical Institute, Scientific Research Centre of the Slovenian Academy of Sciences and Arts, Slovenia

Abstract: The Revised Generic Occupational Stress Index questionnaire (RG-OSI) employs the cognitive ergonomics approach that quantifies the burden of stressors on cognitive resources of the employee. The model is structured as a 2-dimensional matrix, where each element is scored from 0 to 2 (sometimes with intermediate values of 0.5, 1.5 or 1.75) as a combination of various items based on multiple criteria. Due to uncommon scoring system of the questionnaire, our study aimed to explore the appropriateness of the existing scoring system and to get some information on validity of the scale on a Slovenian sample. The questionnaire has been applied on 349 Slovenian employees from different occupational groups and the data were analysed by means of correspondence analysis, classical reliability and item analysis and item response theory analysis. The results of correspondence analysis demonstrate that the response categories on individual variables are not always ordered. Furthermore, we conducted reliability analysis for scales, developed short versions of the scales, and obtained some preliminary information on their validity. The current study provides evidence that the described original scoring system in psychological measures may not be appropriate from the psychometric viewpoint.

Keywords: Occupational Stress Index, test construction, correspondence analysis, item response analysis, reliability

Psihometrične značilnosti slovenske priredbe revidirane verzije generičnega vprašalnika Indeks poklicnega stresa (RG-OSI)

Nataša Sedlar^{1*}, Gregor Sočan² in Lilijana Šprah³

¹Nacionalni inštitut za javno zdravje, Ljubljana

²Oddelek za psihologijo, Filozofska fakulteta, Univerza v Ljubljani

³Znanstvenoraziskovalni center Slovenske akademije znanosti in umetnosti, Ljubljana

Povzetek: Revidirana verzija generičnega vprašalnika Indeks poklicnega stresa (Occupational Stress Index; RG-OSI) temelji na kognitivno-ergonomskem pristopu, ki poklicni stres analizira z vidika obremenitev na kognitivne procese zaposlenih. Model je strukturiran kot dvodimenzionalna matrika, znotraj katere je vsak element točkovan z vrednostmi od 0 do 2 (včasih z vmesnimi vrednostmi 0,5, 1,5 ali 1,75), ki so določene kot nelinearne kombinacije več kriterijev iz različnih postavk. Ker uporabljen način točkovanja v psihološkem merjenju ni pogost, smo v naši študiji želeli preveriti njegovo ustreznost in zbrati nekaj informacij o veljavnosti slovenske priredbe vprašalnika. Vprašalnik je izpolnilo 349 slovenskih zaposlenih iz različnih poklicnih skupin. Podatke smo analizirali s korespondenčno analizo, klasično analizo zanesljivosti in postavk, ter analizo TOP (teorije odgovora na postavko). Rezultati korespondenčne analize kažejo, da se kategorije odgovorov na posameznih spremenljivkah pogosto ne razvrščajo po velikosti. V nadaljevanju smo na podlagi analize zanesljivosti razvili kratko verzijo vprašalnika in zbrali nekaj podatkov o njeni veljavnosti. Raziskava kaže, da psihometrične značilnosti opisane načina točkovanja v psihološkem merjenju niso najbolj ustrezne.

Ključne besede: Vprašalnik Indeks poklicnega stresa, sestava testa, korespondenčna analiza, analiza postavk, zanesljivost

*Naslov/Address: Nataša Sedlar, Nacionalni inštitut za javno zdravje, Trubarjeva 2, 1000 Ljubljana, e-pošta: natasa.sedlar@gmail.com

Changes in content and organisation of work in recent decades have resulted in an increasing exposure to psychosocial risks in the workplace that can considerably impair workers' and organisational health. Several studies (EU-OSHA, 2010; Parent-Thirion, Fernández Macías, Hurley, & Vermeylen, 2005; Parent-Thirion et al., 2010) showed that exposure to psychosocial risk factors at work may result in a state of work-related stress that has negative impact on employees, organizations and national economies. Longitudinal studies and systematic reviews (Niedhammer, Tek, Starke, & Siegrist, 2004; Salavec et al. 2010; van Stolk, Staetsky, Hassan, & Kim, 2012) have indicated its association with heart disease, anxiety, depression, and musculoskeletal disorders. Moreover, there is strong and consistent evidence that high job demands, low control, and effort-reward imbalance present risk factors for mental and physical health problems (Halbesleben & Buckley, 2004; Siegrist, 2002).

Due to its negative impact on employees and work organisations, work-related stress has been widely studied. Most commonly used models for studying the effects of psychosocial work factors on workers' health are the demand-control model (JDC or job strain; Johnson & Hall, 1988; Karasek, Baker, Marxer, Ahlbom, & Theorell, 1981) and Effort-Reward Imbalance model (ERI; Siegrist, 2002). The JDC model emphasizes the extrinsic and situational components of the work (job demands, control/decision latitude and support at work), while the ERI model incorporates the intrinsic characteristics of an individual (motivation, commitment to the work). According to the JDC model exposure to high levels of psychological demands and low levels of social support and job control are associated with more negative health outcomes. The ERI model, on the other hand, proposes that high efforts of employees (external demands or internal motivations), if not matched with high rewards (economic, recognition, promotion prospects, job security, etc.), may as well lead to strain and negative health outcomes of the employees. Due to the popularity of the described models occupational stress is mostly assessed using their corresponding instruments (e.g., Job Content Questionnaire, Effort-Reward Imbalance questionnaire).

Complementary to sociological models of work-related stress, but less frequently used, are models arising from cognitive ergonomics approach. Cognitive ergonomics is most commonly defined as a discipline that focuses on mental processes such as perception, memory, information processing, reasoning, and motor response as they affect interactions among humans and other elements of a system (Hollnagel 2003; Vicente, 1999). One of the current models, developed from a cognitive ergonomics, is Occupational Stress Index (OSI) model (Belkić, 2003). It is based on Welford's information processing model (Welford, 1968) and incorporates risk factors for work-related stress at different levels of information transmission: sensory input, decision process and action. Therefore, factors as nature and temporal density of incoming information as well as complexity, completeness and coherence of the processed information are included to assess the burden

of work processes on the central nervous system of the employee. The model includes risk factors arising from different levels of work environment; from task-level, work schedule, physical and chemical factors to broader organizational factors. Taking this theoretical approach Belkić and Savić (2008; Belkić, 2003) developed the Revised Generic OSI questionnaire (RG-OSI).

The Revised Generic OSI questionnaire

The generic form of the questionnaire is applicable to workers of any occupational profile and allows between-group analyses. It can also be taken as a starting point for the development of occupation-specific OSI questionnaires (e.g., Belkić, Emdad, & Theorell, 1998; Belkić & Nedić, 2007). Following a cognitive ergonomics perspective work-related stress is evaluated in terms of demands of the work on mental resources of the employee.

The OSI model is structured as a 2-dimensional matrix. Seven stress aspects (underload, high demand, strictness, extrinsic time pressure, aversive/noxious exposures, avoidance/symbolic aversiveness, conflict/uncertainty) are combined with 4 levels of information transmission (input, central decision making, task performance, and general level; Figure 1). First three levels of information transmission represent the basic cognitive ergonomic processes as described by Welford (1968), whereas 'general' level is added to include the elements related to broader work context.

Each stress aspect therefore includes risk factors from all three levels of information transmission and some general risk factor as presented in Table 1. For example, underload stress aspect includes 'low frequency of incoming signals' and 'no communication needed' on input level, 'automatic decisions' on central decision-making level, 'homogenous tasks' and 'simple tasks' on task performance level and 'inadequate pay' and 'no chances for upgrade' on general level. Summations by levels of information transmission and by stress aspects can be made. The sum of all scores represents an attempt to quantify the overall burden of working condition on an employee.

Technically, each risk factor is defined as a pair of coordinates, defined by the type of stress and the level of information transmission. It is scored according to an intricate scoring system; from 0 (absence) to 2 (strongly present), sometimes with intermediate values of 0.5, 1.5 or 1.75, where scores are related to (typically non-linear) combinations of multiple criteria. Number of scoring categories for each risk factor and corresponding scoring rules were based on subjective judgment of the authors. Detailed instructions about which items need to be taken into account to get scores for each element in the matrix are provided in the scoring sheet. For example, the risk factor 'automatic decision making' (placed on the central decision-making level of stress aspect underload) can be scored with 0, 1 or 2 points. Zero points refers to the absence of the risk factor and is obtained when

Table 1. *The Occupational Stress Index (OSI), Version 2003, (Belkić and Nedić, 2007, p. 63)*

Levels of Information Transmission	<i>Underload</i>	<i>High Demand</i>	<i>Strictness</i>	<i>Extrinsic Time Pressure</i>	<i>Aversiveness (Noxious Exposures)</i>	<i>Avoidance (Symbolic Aversiveness)</i>	<i>Conflict / Uncertainty</i>
Input	*Homogeneous signals *Low frequency of incoming signals *Works alone – without need to communication	*Several info. sources *Heterogeneous information *Heavy burden on visual system *High frequency of incoming signals *Three sensory modalities *Communication essential	*Strict requirements for signal detection	*No control over speed of incoming signals	*Glare *Noise	*High level of attention (serious consequences of a momentary lapse) *Visually-disturbing scenes *Listens to emotionally-disturbing occurrences	*Signal / Noise conflict *Signal / Signal conflict
Central Decision-Making	*Decisions automatic from input	*Complex decisions *Complicated decisions *Decisions affect work of others *Rapid decision-making	*Strict problem-solving strategy *Strictly defined correct decisions	*Decision cannot be postponed		*Serious consequences of a wrong decision	*Missing information needed for decision *Contradictory information *Unexpected events change work plan
Output / Task Execution	*Homogeneous tasks *Simple tasks *Nothing to do	*Heterogeneous tasks *Simultaneous task performance *Complex tasks *Rapid task performance	*Work must meet a strictly-defined standard	*No control over rate of task performance	*Vibration *Isometric lifting	*Hazardous task performance	*Conflicting demands Task performance hampered by: *Extrinsic problems *Interruptions from people
General	*Fixed pay *Inadequate pay *No chances for upgrade *Lack of recognition of work	*Piece rate work *Long work hours *Holds 2+ jobs *Lack of rest breaks *Night shift/irregular work hours *Lack of paid vacations	*Fixed body position *Confined, window-less workspace *Lack of autonomous work-space *Limited in taking time off from work *Low influence over: Schedule, Policy, tasks, with whom one works	*Speed-up *Deadline pressure	*Cold *Heat *Noxious gases/fumes/dusts	*Work accident *Witness work accident *Suicide occurrence *Work-related litigation/testifying in court *Lack of functioning emergency system	*Emotionally charged work atmosphere *Lack of help with work-related difficulties *Opposition to career advancement *Violation of behaviour norms/abuses of power *No grievance redress *Threat of job loss *Lack of job coherence

the respondent's answers reflect that he is not making automatic decisions; Decisions not automatic (I2 = c, d or e) OR Some supervisory work (I1 = b, c or d). On the contrary, 2 points are ascribed when the content of included items reflects automatic decision making of the respondent; Fully automatic decisions (I2 = a) AND No supervisory work (I1 = a); see Table 2. Despite the fact the RG-OSI is used in organizational research (e.g., Soori, Rahimi, & Mohseni, 2008), there is no available evidence concerning the validity of the scoring rules and the psychometric characteristics of the instrument in general.

Aim of the study

Both the classical test theory approach and the congeneric (factor-analytical) approach to the analysis of

the internal structure of a test imply a (possibly weighted) sum score. The unique scoring system of the RG-OSI questionnaire is therefore not suitable for analysis with psychometric tools, based on the linear model, because the total test scores are not linear composites of the ratings at the most elementary level. Furthermore, the item response theory (IRT) is not optimal either, since IRT—similarly to the previously mentioned approaches—assumes that the test components can be freely combined. Nevertheless, we aimed to explore the appropriateness of the existing scoring system and to estimate the reliability of the scale on a Slovenian sample, using the classical and IRT approach, since we could find no viable non-compensatory model that could be fitted to our data. In order to get some initial assessment of the validity of RG-OSI, its correlations with some negative stress outcomes (intent to turnover, burnout, and work-family conflict) were explored.

Method

Participants

Data collection was carried out within different occupations from public and private sector where several employers have been invited to participate in the study. The approval of the local psychological ethics committee has been obtained prior to the study. After giving informed consent, employees were asked to anonymously complete the questionnaires according to the instructions and return it in a sealed envelope.

The participants ($N = 141$) were employees of four work organizations from different sectors (health, construction, industrial work) that took part in the project »The Support Programme for Employers and Employees for Reducing Work-related Stress and Its Adverse Effects«. The rest of the data were collected on employees of the Slovenian Association of Free Trade Unions ($N = 33$), employees of various police directorates across Slovenia ($N = 83$), and on an opportunity sample of Slovenian employees in different occupations ($N = 92$). The sample ($N = 349$) predominantly consisted of employees from government, public administration and defence (34.6%), health care and social work (20.8%), industry or manufacturing (18.3%), construction (8.7%), and education (5.1%). According to Things-Data-People taxonomy (Fine & Cronshaw, 1999) the majority of participants worked primarily with people (39.7%), 36.6% worked primarily with things and 20.8% worked primarily with information. 170 participants (48.7%) were male. More than a third of the sample was 31 to 40 years old (35.1%), 26.4% of the participants were 41 to 50 years old, 22.6% 20 to 30 years old, 15.9% more than 50 years old, and 8.1% less than 20 years old. Regarding education, most of the participants completed either high school (35.5%), university (28.5%), vocational (7.6%) or higher vocational school (11.8%). The mean working experience was 14.7 years ($SD = 12.2$), the mean organizational tenure was 9.6 years ($SD = 8.9$) and most of the participants (90.2) worked under long-term contract.

Instruments

Workplace stress was assessed by the Revised Generic OSI questionnaire (RG-OSI, Belkić & Savić, 2008) that consists of 65 items. It measures risk factors at different levels of information transmission (input, central decision making, task performance and general level) that cover 7 stress aspects: Underload, High Demand, Strictness, Extrinsic Time Pressure, Aversive/Noxious Exposures, Avoidance/Symbolic Aversiveness, Conflict/Uncertainty. Higher score on a particular stress aspect and higher total OSI score indicate higher level of work-related demands on mental resources of the employee. Detailed information about the instrument is provided in the introductory section of the article.

Intent-to-turnover was measured with two items, adapted from existing questionnaires measuring an

employee's desire to remain with the current employer (Konovsky & Cropanzano, 1991; Scott, Bishop, & Chen, 2003): 'I often think of quitting my current job', 'It is very possible for me to leave for another company next year'.

Burnout was measured with The Oldenburg Burnout Inventory (OLBI; Demerouti, Bakker, Nachreiner, & Schaufeli, 2001). Items measuring two dimensions of burnout exhaustion and disengagement are scored on a four-point scale from strongly agree (1) to strongly disagree (4). Exhaustion subscale refers to general feelings of emptiness, overtaxing from work, a strong need for rest, and a state of physical exhaustion. Disengagement subscale refers to distancing oneself from the object and the content of one's work and to negative, cynical attitudes and behaviours toward one's work in general. Each subscale includes four items that are positively framed and four items that are negatively framed. The reliability of the scales in Slovenian validation study of the questionnaire (Sedlar, Šprah, Tement, & Sočan, submitted) was moderate ($\alpha = .73$ for Exhaustion scale and $\alpha = .71$ for Disengagement scale). Higher reliability ($\alpha = .88$) was obtained for scales consisting only of negatively framed items. The »negative items only« model also fitted the data better than other models, tested by means of confirmatory factor analysis. Therefore, only this part of the original questionnaire was used in our study.

Work-family conflict was measured with the Slovenian version of Work-family conflict scales (Carlson, Kacmar, & Williams, 2000; Tement, Korunka, & Pfifer, 2010). For the purpose of this study, we used only three subscales measuring more difficult participation (lack of time, strain or incompatible behaviour) in the family domain because of work responsibilities. Subscales Time-Based Work-Family Conflict and Strain-Based Work-Family Conflict refer to work and family competing for a person's time or energy, respectively, while subscale Behaviour-Based Work-Family Conflict refers to incompatible behaviours between these two domains. Each dimension consists of three items with a five-point response scale (1 = strongly disagree; 5 = strongly agree).

Procedure

The process of translation and adaptation of the RG-OSI instrument

Implementation of the Slovenian RG-OSI instrument underwent the following steps:

1. *Forward translation*
Two psychologists, experts in the field of occupational psychology with advanced knowledge of English language translated the original version of the RG-OSI instrument into the Slovenian language.
2. *Expert panel on the suitability of the EN-SLO translation*
Four experts (two general psychologists, one psychologist specialized in health and one psychologist

specialized in occupational psychology, all of them experienced in instrument development and translations) reviewed discrepancies between English version of the RG-OSI and Slovenian translations, as well as suggested alternatives. They ensured that proposed alternatives were conceptually and culturally equivalent to the original.

3. Pre-testing of the RG-OSI instrument and cognitive interviewing

Pre-testing of the instrument was carried out in the group of 11 respondents. They represented males and females from all age groups (18 years of age and older), with different socioeconomic backgrounds and different occupations. The interviews were conducted by experienced interviewers. Each respondent individually filled out the RG-OSI questionnaire and was afterwards systematically debriefed. During the interview, they were asked the following set of questions: 1) Do you find the question understandable? 2) Could you repeat it in your own words? 3) Did you have any problems with answering it? 4) Were there any words you did not understand or any word or expression that you find unacceptable or offensive?

4. Final version and back-translation

The final version of the instrument in the Slovenian language has been created after all stages described above and translated back to the English by an independent translator, whose mother tongue was English and who had no previous knowledge of the questionnaire. The English back-translation of the Slovenian version of the RG-OSI instrument has been introduced to the author of the original RG-OSI questionnaire, who approved modifications and its use in Slovenia. Final version of a translated questionnaire closely resembled the English version.

Analyses

Inspection of the data revealed that all variables have at least one missing value and the maximum number of missing values was 19. On average there were 1.9% missing values per variable. In order to avoid the loss of these data, missing data were estimated. Although Little's MCAR test (Little & Rubin, 1989) was significant ($p = .01$), we see no substantive reasons to expect the presence of Missing Not At Random (MNAR) missing mechanisms. Missing data were imputed using the Expectation-Maximization (EM) algorithm that estimates missing values by an iterative process and has been demonstrated to be an effective method of dealing with missing data (Graham, 2009). All analyses were conducted using a total of 349 participants.

SPSS Statistics 21.0 was used to perform all the statistical analyses, except for item response analysis that was done in IRTPRO (Cai, Thissen, & du Toit, 2011). A correspondence analysis was conducted to check whether multivariate categorical responses are ordered (i.e., scaled on at least an ordinal scale). To evaluate the ability of each

item to discriminate between levels of different stress aspects the classical and the item response analysis were conducted. The reliability of the scales was estimated with internal consistency coefficients (α). Items with lower discrimination power were eliminated and the correlations of both the obtained shorter and the original scales with selected variables were compared.

Results

This section consists of three parts. We begin with the presentation of the scoring system for one of the scales, followed by evaluation of the scoring system by means of correspondence analysis and IRT analysis. Finally, we present the results of the reliability analysis for variant scales and their correlations with selected variables.

Evaluation of the scoring system (subscale Underload)

For the purpose of this article, we will illustrate the evaluation of the scoring system of the first dimension Underload. Table 2 presents the structure of the dimension according to the cognitive ergonomic theory. The total score is obtained as a sum of risk factors on three levels of information transmission (input, central decision making, task performance) and general risk factors. Information about items that need to be taken into account to get scores for each risk factor and corresponding scoring rules are presented in the scoring sheet (see Table 3 for a demonstration).

Table 2. Structure of the Underload dimension on different levels of information transmission

Homogeneous information = IU1
Low frequency of incoming signal= IU2
No need for communication= IU3
Input Underload Total (IUT) = IU1 + IU2 + IU3
Automatic decision-making= CUT
Central Underload Total (CUT) = CUT
Homogeneous tasks= OU1
Simple tasks= OU2
Nothing to do= OU3
Output Underload Total (OUT) = OU1 + OU2 + OU3
Fixed pay= GU1
Inadequate pay= GU2
Lack of promotion prospects= GU3
Lack of recognition of good work= GU4
General Underload Total (GUT) = GU1 + GU2 + GU3 + GU4
Total Underload Score= IUT + CUT + OUT + GUT

Number of variables on different levels of information transmission differs; Input and Output level consist of three risk factors, Central decision making of one risk factor, and General level of four risk factors. Total Underload score is calculated as a sum of all subscales.

Some of the risk factors are scored with 0 and 2 (IU2, IU3, OU2, GU1), some with 0, 1 and 2 (IU1, CUT, GU3), while risk factors (OU1, OU3, GU2, GU4) also include values of 0.5 and 1.5. Scoring instructions in the right column show that some scores are obtained from responses on a single item (OU1, OU3, GU1-GU4) and some are related to a (non-linear) combinations of multiple criteria (IU1-IU3, CUT, OU1, OU2), as exemplified in the introduction.

In the first step, we checked whether the possible item score values are empirically ordered (for instance, whether a score 0.5 statistically implies a lower level of the measured trait than a score of 1). We used a multiple correspondence analysis for this purpose. The correspondence analysis is primarily a (multivariate) descriptive data analytic technique (in which case it is analogous to principal component analysis), but it can also be used as a scaling method (Greenacre, 2007; McDonald, 1999).

The largest amount of variability in the data points is captured by the first dimension; in our example, the first dimension explained 20% of inertia and had a moderate reliability estimate ($\alpha = 0.61$). Frequency distributions of

Table 3. Scoring sheet for the Underload dimension

	0	Heterogeneous information (H3=a or b (no monotonous tasks)) OR (H1=several tasks) OR (J2-9=interacts with persons or several machines)
IU1	1	Moderately homogeneous information (H3=c or d) AND [(H1= a few tasks) OR (J2-9=limited interactions with persons / few machines)]
	2	Maximally homogeneous information (H3=d (monotonous tasks)) AND (H1=very few, simple tasks) AND (J2-9=minimal interactions with persons/ few machines)
IU2	0	Controls speed him or herself (F3=a or b AND if J3=yes, controls speed of devices) OR >1 new signal/minute (overall assessment, see especially J3-9)
	2	Doesn't control speed (F3=c or d, OR J3=yes and doesn't control speed of devices) AND <1 new signal per minute (overall assessment, see especially J1-9)
IU3	0	J1=b or c (works with others)(Verify with J4-9)
	2	J1=a (No need for communication with others)((Verify with J4-9)
CUT	0	Decisions not automatic (I2=c, d or e) OR some supervisory work (I1=b, c or d)
	1	Fairly automatic decisions but judgment required (I2=b) AND No supervisory work (I1=a)
	2	Fully automatic decisions (I2=a) AND no supervisory work (I1=a)
OU1	0	H3=a or b (Heterogeneous tasks)
	1	H3=c (Fairly homogeneous tasks)
	1.5	H3=d (Homogeneous tasks) AND H1=three or more tasks
	2	H3=d (Homogeneous tasks) AND H1=two or fewer tasks
OU2	0	Several steps (assessed from H1 & H2)
	2	Few steps in tasks (assessed from H1 & H2)
OU3	0	H4=a (Always something to do)
	0.5	H4=b (Rarely nothing to do)
	1	H4=c (Occasionally nothing to do)
	2	H4=d (Frequently nothing to do)
GU1	0	C1=a or b (Based upon amount of individual or group work)
	2	C1=c (Fixed pay)
GU2	0	C2=a (Covers substantially more than basic needs)
	0.5	C2=b (Covers a bit more than basic needs)
	1	C2=c (Just barely covers expenses)
	2	C2=d (Totally inadequate)
GU3	0	C3=Yes + a, b, or c (There are possibilities to upgrade job title or advance one's career, and no active opposition)
	1	C3=Yes + d (There are upgrade possibilities, but active opposition)
	2	C3=No
GU4	0	C5=a (Definitely yes)
	0.5	C5=b (Yes, to some extent)
	1.5	C5=c (Not very much)
	2	C5=d (Not at all)

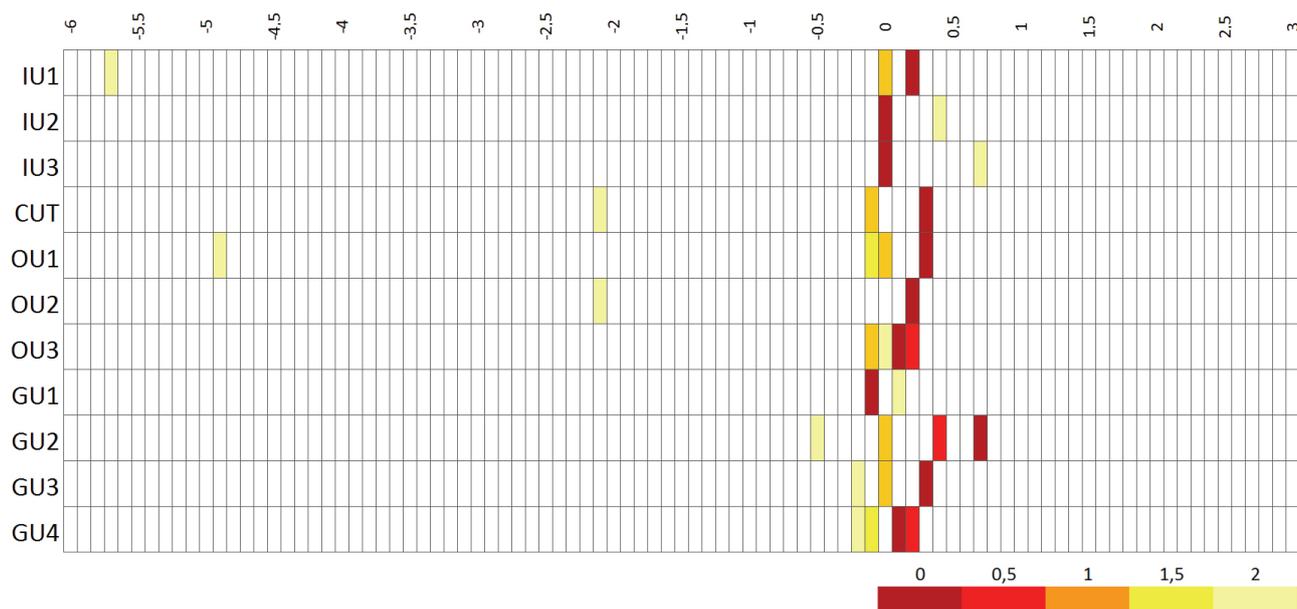


Figure 1. Graphical presentation of response scale values.

the variables (not shown in the paper) showed skewed distribution of categorical answers for the first few variables included into the Underload scale (IU1-OU2), because the majority of the participants answered with 0. To illustrate the relative ranking and ordering of centroid coordinates for response categories a graphical presentation was made (Figure 1). Centroid coordinates for response categories of a certain variable should be placed on an ordinal scale if the response categories were accurately measuring burden of a certain risk factor (higher number of points on a variable means higher burden than lower number of points). Colours in rows should therefore be arranged from bright to dark or vice versa. As we can see, this was not the case for variables IU1, OU3 and GU4, where the intermediate values were not ordered as expected. We can also note large differences in the spread of scaled positions of the score values. Items with smaller spread are more likely to have low discrimination indices than items with larger spread, which will be demonstrated in the follow-up.

In the next step, item response analysis using the IRTPRO programme (Cai et al., 2011) was performed to evaluate Underload scale variables. Analysis was conducted using all 11 variables related to all four levels of information transmission, because the number of variables on separate levels of information transmission was too small.

Item response theory (IRT) relates characteristics of items and characteristics of individuals to the probability of choosing each of the response categories. This probabilistic relationship is mathematically defined as a nonlinear regression of the probability of choosing an item response category on a latent trait (item response function). The two-parameter logistic model (2 PL) was used for dichotomously scored variables. For variables with more than two response options, the graded response

model was used, which is a generalization of the 2 PL model for polytomous items (for more information on the models used and item response theory in general, see for instance Embretson & Reise, 2000).

Table 4 presents the item statistics for each item. The item discrimination indices, item thresholds, and the test of model fit are presented. The discrimination index should always be positive, and values above 2 generally indicate a very good discrimination. The threshold (difficulty) index of category j is the value of the measured trait (in our case, Underload), at which the probability of responding with category j (or higher) is 50%. For dichotomous items, this simplifies to the trait value where the positive and the negative response are equally likely.

Variables IU1 and OU1 had very high discrimination, while the other variables discriminated moderately or non-adequately between respondents. The chi-square goodness-of-fit test indicated a lack of fit for domains IU1, OU1, GU2, GU3; for the remaining majority of the items, the fit seemed to be satisfactory. Since the domains are classified into levels of transmission, local dependence of items within a level could be expected; however, the analysis of residuals (not presented here) did not reveal any problems in this respect. The highest absolute residual value was 2.9, which is well below the value of 10, which Cai et al. (2011) consider to be high.

Item information curves showed that items IU1 and OU1 provided almost all information (see Figure 2), while the other items had little information value (see Figure 3 for a sample graph). The results of item response analysis also showed that the intermediate values of response categories were problematic from the discrimination point of view (see Figure 4). Figure 4 shows that in case of domain GU3, the intermediate item score category 1 was practically useless: its category response function was almost flat, which means that this category did not play

Table 4. Item analysis statistics for the Underload dimension

Variable	Discrimination (a)	Difficulty (b_1)		χ^2	df	p
2PL model						
IU2	6.23	0.66		2.74	3	0.43
IU3*	-70.38	-0.08				
OU2	3.55	0.92		10.93	8	0.21
GU1	14.94	-0.14		8.47	9	0.49
Graded model						
		Threshold (b_1)	Threshold (b_2)	Threshold (b_3)		
IU1	0.78	1.65	2.08		20.37	6
CUT	0.21	2.19	3.29		24.76	18
OU1	0.85	0.63	1.56	1.99	48.94	18
OU3	0.14	-2.20	0.04	1.66	31.45	28
GU2	0.14	-8.23	-0.98	3.74	40.36	22
GU3	0.15	-1.01	-0.68		30.00	15
GU4	0.14	-3.51	0.91	5.60	26.37	26

*IRTPRO could not perform the test of fit.

any role in discriminating between persons. In general, category response curves typically emerge when low discrimination is combined with relatively small distances between category thresholds (see Table 4).

We might also note that extreme difficulty estimates of some dichotomous items (in this case, IU3 and GU1) are related to their low discrimination: when the slope of the item characteristic curve is low, a small variation of the slope results in a large difference in the threshold value, which is also reflected in large standard errors for such parameters (standard errors are not shown in Table 4; for instance, the standard error for difficulty parameter of domain IU3 was about 1066). Such values should be therefore considered with great reservation.

We do not discuss the results for the remaining dimensions in details; we only note that the situation was quite similar. There were many items with either low or implausibly high estimate of the discrimination parameter (the latter often in combination with implausibly high threshold values). Furthermore, the analysis of residuals revealed slight local dependency problems (i.e., one or two high local dependence (LD) values larger than 10) with two scales and severe local dependency problems in three scales. For more detailed information see the Appendix.

Reliability analysis

A reliability analysis of all dimensions was conducted, first on the level of subscales (i.e., Information transmission level of stress aspects). Only the results for the dimension Underload are presented to illustrate the procedure and computations. The values of the reliability coefficients, corrected item-total correlations and reliability coefficients if item deleted are presented in Table 5. Reliability coefficients were very low ($\alpha = 0.20-0.27$), indicating a limited psychometric usefulness of the scale that could be due to poor interrelatedness between items or too heterogeneous constructs. Therefore, the variables with the lowest item-total correlations (0.00-0.09) were deleted and the analysis was conducted on the remaining seven variables.

In the second round of item elimination (Table 6), items IU2 and IU3 were deleted. The short Underload scale therefore consisted of 5 variables, but still had a relatively low reliability ($\alpha = .40$).

After we have conducted the reliability analysis, all the IU items were eliminated. Hence the Underload dimension of the short RG-OSI scale in the end had no input items.

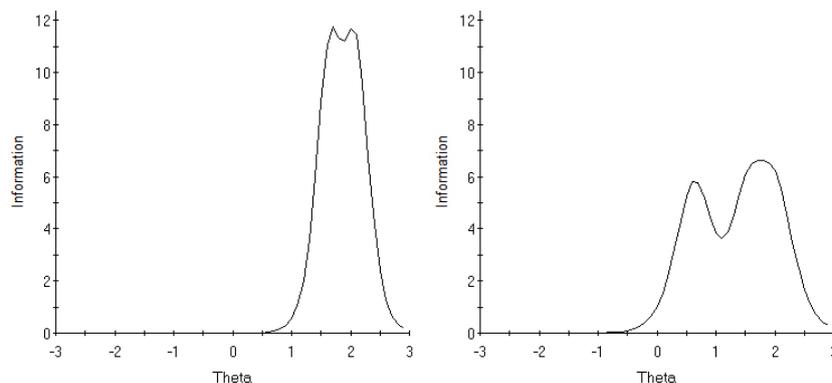


Figure 2. Item information curves for the domains IU1 (on the left) and OU1 (on the right).

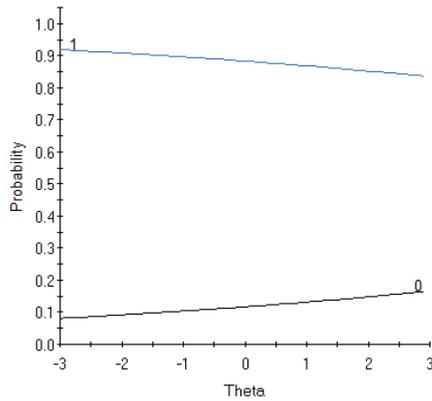


Figure 3. Item characteristic curve for the domain GU1. The graph shows the probability of choosing an item response category on a latent trait theta.

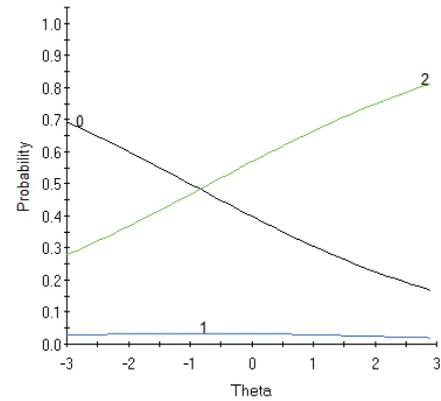


Figure 4. Item characteristic curve for the domain GU3. The graph shows the probability of choosing an item response category on a latent trait theta.

In a similar manner items were sequentially removed from other dimensions to maximise reliability. In Table 7 it can be seen that it was necessary to eliminate only a few items from the dimensions Extrinsic Time Pressure (one item), Aversive/Noxious Exposures (two items), Avoidance/Symbolic Aversiveness and Conflict/Uncertainty (three items). More items had to be removed from dimensions Strictness (five), Underload (seven) and High Demand (nine). After eliminating redundancies Extrinsic Time Pressure dimension included the lowest number of items (five), whereas dimension Conflict/Uncertainty consisted of the highest number of items (twelve). Reliabilities of scales after two rounds of item elimination were ranging from relatively low (Extrinsic Time Pressure $\alpha = .50$; Strictness $\alpha = .66$) to moderate

(Aversive/Noxious Exposures $\alpha = .70$; High Demands $\alpha = .71$; Avoidance/Symbolic Aversiveness $\alpha = .80$; Conflict/Uncertainty $\alpha = .83$). Similar to Underload dimension, Extrinsic Time Pressure dimension and Strictness dimensions of the short RG-OSI scale in the end had no input items, whereas the latter also lacked output items.

Evaluation of the shorter version

The aim of the construct validation is to embed a purported measure of a construct in a nomological network, that is, to establish its relation to other variables (Cronbach & Meehl, 1955). Although the construct validation of RG-OSI was not the principal aim of this study, we tried to gather some preliminary information

Table 5. Reliability estimates for the subscales of Underload dimension

	α	r_{jt}	$\alpha_{(i)}$
Homogeneous information = IU1*		.08	.44
Low frequency of incoming signal = IU2		.23	-.01
No need for communication = IU3		.23	.19
Input Underload Total (IUT) = IU1 + IU2 + IU3	.27		
Automatic decision-making = CUT			
Central Underload Total (CUT) = CUT			
Homogeneous tasks = OU1		.21	-.07
Simple tasks = OU2*		.04	.26
Nothing to do = OU3*		.09	.23
Output Underload Total (OUT) = OU1 + OU2 + OU3	.20		
Fixed pay = GU1*		.00	.28
Inadequate pay = GU2		.19	.08
Lack of promotion prospects = GU3		.13	.12
Lack of recognition of good work = GU4		.11	.15
General Underload Total (GUT) = GU1 + GU2 + GU3 + GU4	.20		

Note. r_{jt} = item discrimination (corrected item-total correlation); $\alpha_{(i)}$ = coefficient alpha if item deleted

* eliminated items

Table 6. Classical item statistics for Underload dimension after first round of item elimination

	r_{ji}	$\alpha_{(i)}$
Low frequency of incoming signal = IU2*	.02	.41
No need for communication = IU3 *	.06	.41
Automatic decision-making = CUT	.25	.32
Homogeneous tasks = OU1	.27	.32
Inadequate pay = GU2	.29	.31
Lack of promotion prospects = GU3	.20	.39
Lack of recognition of good work = GU4	.18	.36

Note. r_{ji} = item discrimination (corrected item-total correlation); $\alpha_{(i)}$ = coefficient alpha if item deleted
* eliminated items

on the relations between RG-OSI scales and some related variables. Specifically, we tested whether the participants implicitly differentiated between stress aspects measured by RG-OSI and different constructs (fluctuation, burnout and work-family conflict). Original and short RG-OSI scales were correlated with the respective measures. The results are presented in the Table 8.

The low correlations between the OSI scales and the correlates imply that the participants differentiated between different constructs, but on a limited range. Moreover, the newly constructed shorter scales seemed to have almost identical correlations with measures of similar variables than the original scales.

Discussion

We evaluated the basic psychometric structural characteristics of the Slovenian translation of the RG-OSI. Due to the less commonly used cognitive-ergonomics approach that tries to quantify the burden of work-related stress on mental resources of the employee, the questionnaire seemed to be a useful alternative to the existing questionnaires measuring work-related stress.

However, one of the major restraints seemed to be an intricate scoring system, where each variable score is obtained as a combination of multiple criteria. Both correspondence and item-response analysis indicated that this approach is of limited psychometric value. Centroid coordinates indicated that response categories are not ordered, which may put the predetermined scoring system established by original authors under question (example in Table 3). Intermediate values seem to be especially problematic, also from the discrimination point of view. The examination of item characteristic curves indicated that most of the information from the examined dimension is provided in 2 items (we remind the reader that the maximum information of an item is related to its discrimination), while other items had considerably smaller information value and hence smaller discrimination power. The major reason for this most probably resides

Table 7. Number of items for original and short RG-OSI scale and coefficient α 's for short RG-OSI scale

	Number of items RG-OSI		α short
	original	short	
Underload	11	5	.40
High demands	20	11	.71
Strictness	12	6	.66
Extrinsic time pressure	5	4	.50
Aversive/Noxious exposures	7	6	.70
Avoidance/Symbolic aversiveness	10	9	.80
Conflict/Uncertainty	15	12	.83

in the construct's definition that is too heterogeneous to be conceptualized as a unidimensional latent trait. The fact that only two out of seven subscales appeared to be free of local dependency indicates a possible need for a multidimensional representation. Unfortunately, the instrument in its present form does not include enough items to enable such modeling.

Based on the results of the reliability analysis (Table 5), items with low discrimination power were eliminated. As a result of two rounds of elimination process, the final scale consisted of five instead of eleven variables. Similar procedure was conducted for all subscales. Correlations between shortened scales and related variables constructs (fluctuation, burnout and work-family conflict) were compared to correlations between the original RG-OSI scale and before mentioned external correlates, showing no major differences (Table 8). We would expect that better internal structure of the short RG-OSI (according to item response analysis and reliability analysis) would result in higher correlations with other variables. It is possible that this was not the case because some important contents were eliminated from the questionnaire (e.g., Underload dimension has no input underload items), which seems problematic from cognitive-ergonomics perspective. In order to assess the overall burden of work circumstanced on the employee, the cognitive-ergonomics approach presumes that demands on mental resources of the employee include different levels of information transmission: sensory input, decision process and action input level. Absence of any of these aspects could therefore severely impair the validity of the questionnaire. On the other hand, these results may be an instance of a well-known phenomenon: increasing internal consistency may decrease the correlations with complex criteria (McDonald, 1999). Still, we believe that finding scales with a sound internal structure should always be the first step in a test construction; in the second step, such scales may be combined to achieve higher predictive power.

In general, the results of all internal structure analyses showed weak internal structure, which was reflected in the fact that many items had low spread of scaled values in

Table 8. Kendall's tau correlations of the original and short scales with related constructs

Dimension	ITQ think	ITQ next year	OLBI neg	WFC time	WFC strain	WFC behaviour
Underload total	.16**	.16**	.25**	.17**	.07	.01
Underload short	.18**	.17**	.24**	.20**	.06	-.01
High demand total	.03	.05	.00	.11*	.12**	.08*
High demand short	-.03	.00	-.07	.02	.09	.07
Strictness total	.14**	.10*	.24**	.25**	.06	.01
Strictness short	.14**	.10*	.21**	.22**	.03	-.02
Extrinsic time pressure total	.18**	.10*	.11**	.19**	.12**	.04
Extrinsic time pressure short	.18**	.10*	.11**	.18**	.12**	.04
Noxious exposure total	.05	.06	.18**	.12**	.08*	-.04
Noxious exposure short	.05	.06	.18**	.12**	.08*	-.03
Symbolic aversiveness total	.09*	.06	.15**	.17**	.15**	.06
Symbolic aversiveness short	.08*	.06	.14**	.16**	.13**	.06
Conflict total	.21**	.15**	.23**	.32**	.25**	.19**
Conflict short	.21**	.15**	.22**	.31**	.25**	.17**

Notes. ITQ think - think of quitting job; ITQ next year - very possible to change employer next year; OLBI neg - negative items for measuring burnout; WFC time - Time-based work-family conflict; WFC strain - Strain-based work-family conflict; WFC behaviour - Behaviour-based work-family conflict.

* $p < .05$; ** $p < .01$

multiple correspondence analysis and low discrimination indices (both classical and IRT ones). Since the items in the RG-OSI are relatively clearly presented, the answer should probably be looked for in unclear wording or similar item deficiencies. It seems that the major reason for the weak structure resides in the constructs' definitions. In other words, the measured constructs may be too heterogeneous to be conceptualized as latent traits. Of course, this makes the interpretation of the proposed dimension scores unclear. It is a matter of further research to determine whether a more suitable psychometric model can be found for RG-OSI, or should it be abandoned altogether.

Limitations of the study

First potential drawback concerns a rather specific sample, which has not been randomly selected from the full range of possible occupations. Our sample was predominantly restricted to employees of the government, public administration and defence, industry or manufacturing, construction and education. Moreover, it was overrepresented by employees from 31 to 50 years old, and employees with either completed high school or university. However, establishing norms was not the aim of this study. Although the structure of the sample did not perfectly reflect the structure of the Slovene population of employees, we are convinced that our sample was sufficiently heterogeneous to rule out the interpretation of the low reliability as a consequence of a highly homogeneous sample. Another potential drawback is the reliance on self-report and very long questionnaires containing many items (e.g., RG-OSI consists of 65 items), which could affect concentration and motivation of participants.

Future research and implications for practice

To our knowledge, this is one of the first attempts to assess the quality of the OSI questionnaire in its general form, mainly due to more frequently used occupation specific OSI questionnaires that can be developed from the general model. So far our results showed that the questionnaire in its original and short form has some severe psychometric limitations, which should not be overlooked when interpreting results. The short RG-OSI scales showed better internal characteristics and seemed to behave similarly to the original ones with regard to their relation to some other variables, but lacked some important aspects of work-related demands. Moreover, the use of short RG-OSI would considerably shorten the time needed to complete the questionnaire, but some additional work would be needed to replace the eliminated aspects and improve the validity and criterion correlations. We should also note that reliability estimates of short scales should be taken with some precaution, since they were not determined on a new sample and are therefore overestimated at least to a certain extent.

Hierarchical structure of the OSI model with components at the highest level and more specific attributes nested within components could be better tested using a non-compensatory model with multiple latent dimensions. So far, a multidimensional IRT models are mostly used in the field of cognitive diagnosis, where relatively long tests with heterogeneous items that vary in numbers and types of cognitive operations or skills are normally used (see Embretson & Yang, 2013). Apart from this, there is a lack of non-additive measurement models that would be appropriate for psychological tests consisting of multiple dimensions. Multidimensional models as suggested by

Embretson and Yang (2013) could not be applied to RG-OSI questionnaire because it lacks theoretically and empirically plausible theory of the underlying components and attributes. Our use of mainstream analyses may have therefore done an injustice to the instrument. However, before more suitable models emerge, such results may be viewed as the best available approximations to the optimal evidence about the psychometric quality of the test. Therefore, we cannot recommend a routine use of RG-OSI in applied psychology at this stage.

Conclusions

Evaluation of psychometric adequacy of the Slovenian adaptation of the RG-OSI questionnaire showed that the questionnaire has some severe psychometric limitations regarding its scoring system and the conceptualization of the measured constructs. Shortened scales have been proposed, which would considerably shorten the time needed to complete the questionnaire, but would require further development to cover all the theoretical aspects of the cognitive ergonomics perspective.

Our study also suggested that non-additive measurement models, appropriate for psychological tests consisting of multiple dimensions, should be given further attention.

Acknowledgements. The presented study was a part of »The Support Programme for Employers and Employees for Reducing Work-related Stress and Its Adverse Effects«, co-founded by the European Social Fund, EU (framework of the Operational Programme for Human Resources Development for the period 2007-2013) and by the Slovenian Research Agency, Research Programme Collective Memory and Cultural Dynamics, No. P6-0347.

References

- Belkić, K. (2003). *The Occupational Stress Index: An approach derived from cognitive ergonomics and brain research for clinical practice*. Cambridge, United Kingdom: Cambridge International Science Publishing.
- Belkić, K., Emdad, R. & Theorell, T. (1998). Occupational profile and cardiac risk: Possible mechanisms and implications for professional drivers. *International Journal of Occupational and Environmental Health*, 11, 37–57.
- Belkić, K., & Nedić, O. (2007). Workplace stressors and lifestyle-related cancer risk factors among female physicians: Assessment using the occupational stress index. *Journal of Occupational Health*, 49, 61–71.
- Belkić, K., & Savić, C. (2008). The occupational stress index: An approach derived from cognitive ergonomics applicable to clinical practice. *Scandinavian Journal of Work and Environmental Health*, 6, 169–175.
- Cai, L., Thissen, D., & du Toit, S. H. C. (2011). *IRTPRO for Windows* [Computer software]. Lincolnwood, IL, USA: Scientific Software International.
- Carlson, D. S., Kacmar, K. M., & Williams, L. J. (2000). Construction and initial validation of a multidimensional measure of work-family conflict. *Journal of Vocational Behavior*, 56, 249–276.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302.
- Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands resources model of burnout. *Journal of Applied Psychology*, 86, 499–512.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Embretson, S., & Yang, X. (2013). A multicomponent latent trait model for diagnosis. *Psychometrika*, 78(1), 1–14.
- EU-OSHA - European Agency for Safety and Health at Work (2010). *European Survey of Enterprises on New and Emerging Risks (ESENER): Managing safety and health at work*. Luxembourg, Luxembourg: Office for Official Publications of the European Communities. Retrieved from https://osha.europa.eu/en/publications/reports/esener1_osh_management
- Fine, S. A., & Cronshaw, S. F. (1999). *Functional job analysis: A foundation for human resources management*. Mahwah, NJ, USA: Lawrence Erlbaum Associates.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 6, 549–576.
- Greenacre, M. (2007). *Correspondence analysis in practice* (2nd ed.). Boca Raton, FL, USA: Chapman & Hall, CRC.
- Halbesleben, J. R. B., & Buckley, M. R. (2004). Burnout in organizational life. *Journal of Management*, 30, 859–879.
- Hollnagel, E. (2003). *Handbook of cognitive task design*. Mahwah, NJ, USA: Lawrence Erlbaum Associates.
- Johnson, J. V., & Hall, E. M. (1988). Job strain, workplace social support, and cardiovascular disease: A cross-sectional study of a random sample of the Swedish working population. *American Journal of Public Health*, 78, 1336–1342.
- Karasek, R. A., Baker, D., Marxer, F., Ahlbom, A., & Theorell, T. (1981). Job decision latitude, job demands, and cardiovascular disease: A prospective study of Swedish men. *American Journal of Public Health*, 71, 694–705.
- Konovsky, M. A., & Cropanzano, R. (1991). The perceived fairness of employee drug testing as a predictor of employee attitudes and job performance. *Journal of Applied Psychology*, 76, 698–707.
- Little, R. J. A., & Rubin, D. B. (1989). The analysis of social science data with missing values. *Sociological Methods and Research*, 18, 292–326.

- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ, USA: Lawrence Erlbaum Associates.
- Niedhammer, I., Tek, M. Y., Starke, D., & Siegrist, J. (2004). Effort reward imbalance model and self-reported health: Cross-sectional and prospective findings from the GAZEL cohort. *Social Science & Medicine*, 58, 1531–1541.
- Parent-Thirion, A., Fernández Macías, E., Hurley, J., & Vermeylen, G. (2005). *Fourth European Working Conditions Survey*. Dublin, Ireland: European Foundation for the Improvement of Living and Working Conditions (EUROFOND). Retrieved from <http://www.eurofound.europa.eu/pubdocs/2006/98/en/2/ef0698en.pdf> (8. 4. 2013)
- Parent-Thirion, A., Vermeylen, G., van Houten, G., Lyly-Yrjänäinen, M., Biletta, I., & Cabrita, J. (2010). *Fifth European Working Conditions Survey – 2010*. Dublin, Ireland: European Foundation for the Improvement of Living and Working Conditions (EUROFOND). Retrieved from <http://www.eurofound.europa.eu/pubdocs/2011/82/en/1/EF1182EN.pdf> (8. 4. 2013)
- Salavecz, G., Chandola, T., Pikhart, H., Dragano, N., Siegrist, J., Jöckel, K. H., Erbel, R., Pajak, A., Malyutina, S., Kubinova, R., Marmot, M., Bobak, M., & Kopp, M. (2010). Work stress and health in Western European and post-communist countries: An East-West comparison study. *Journal of Epidemiology and Community Health*, 64(1), 57–62.
- Scott, D., Bishop, J. W., & Chen, X. (2003). An examination of the relationship of employee involvement with job satisfaction, employee cooperation, and intention to quit in U.S. invested enterprise in China. *International Journal of Organizational Analysis*, 11, 3–19.
- Sedlar, N., Šprah, L., Tement, S., & Sočan, G. (submitted 2015). Internal structure of an alternative measure of burnout: Study on the Slovenian adaptation of the Oldenburg Burnout Inventory (OLBI). *Burnout Research*, 2(1), 1–7.
- Siegrist, J. (2002). Effort-reward imbalance at work and health. In P. L. Perrewe, D. C. Ganster (Eds.), *Research in Occupational Stress and Well-being: Vol. 2. Historical and Current Perspectives on Stress and Health* (pp. 261–291). New York, NY, USA: JAI Elsevier.
- Soori, H., Rahimi, M., & Mohseni, H. (2008). Occupational stress and work-related unintentional injuries among Iranian car manufacturing workers. *Eastern Mediterranean Health Journal*, 14(3), 697–703.
- Tement, S., Korunka, C., & Pfifer, A. (2010). Toward the assessment of the work-family interface: Validation of the Slovenian versions of work-family conflict and work-family enrichment scales. *Psihološka obzorja*, 19(3), 53–74.
- Van Stolk, L., Staetsky, E., Hassan, E., & Kim, C. V. (2012). *Management of psychosocial risks at work: An analysis of the findings of the European Survey of Enterprises on New and Emerging Risks (ESENER)*. Luxembourg, Luxembourg: Publications Office of the European Union.
- Vicente, K. J. (1999). *Cognitive work analysis: Towards safe, productive, and healthy computer-based work*. Mahwah, NJ, USA: Lawrence Erlbaum Associates.
- Welford, A. T. (1968). *Fundamentals of skill*. London, United Kingdom: Methuen.

Appendix

Table 9. Item analysis statistics for the remaining dimensions

High demands												
Item	<i>a</i>	<i>s.e.</i>	<i>b</i> ₁	<i>s.e.</i>	<i>b</i> ₂	<i>s.e.</i>	<i>b</i> ₃	<i>s.e.</i>	<i>b</i> ₄	<i>s.e.</i>	<i>b</i> ₅	<i>s.e.</i>
IH1	1.76	0.27	-1.40	0.16								
IH2	1.00	0.22	-3.32	0.60	4.02	0.82						
IH3	-0.12	0.15	x	x	x	x						
IH4	0.72	0.60	6.20	4.65								
IH5	1.02	0.17	-1.58	0.24								
IH6	0.41	0.12	-7.51	2.15	-1.08	0.40						
CH1	x	4.33	0.13	0.02	0.82	0.05						
CH2	x	x	0.12	x	1.23	x						
CH3	1.74	0.18	-0.64	0.10	0.46	0.08	1.33	0.12				
CH4	0.94	0.13	-3.52	0.46	-0.66	0.15	1.48	0.21				
OH1	0.45	0.11	-2.19	0.58	-1.52	0.44	2.02	0.53				
OH2	0.67	0.12	-5.11	0.90	-2.12	0.38	0.97	0.23				
OH3	0.52	0.12	-3.15	0.74	3.35	0.78						
OH4	0.55	0.12	-7.41	1.63	-2.51	0.54	2.43	0.52				
GH1	0.08	0.18	x	x	x	x						
GH2	0.47	0.11	-3.05	0.71	-0.38	0.24	2.01	0.49	6.04	1.39		
GH3	0.01	0.17	x	x								
GH4	0.14	0.10	x	x	-4.64	3.38	-4.11	3.02	5.47	3.95		
GH5	0.36	0.12	0.52	0.32	0.55	0.33	1.86	0.65	2.30	0.78	2.38	0.80
GH6	-0.73	0.13	0.22	0.16	-5.37	0.98	-8.48	1.97				

There were 10 high (>10) LD values indicating local dependency.

Response patterns of six items (CH1, CH2, CH3, OH2, OH3, GH2) had a significant ($p < .05$) lack of fit to their respective model.

Strictness												
Item	<i>a</i>	<i>s.e.</i>	<i>b</i> ₁	<i>s.e.</i>	<i>b</i> ₂	<i>s.e.</i>	<i>b</i> ₃	<i>s.e.</i>	<i>b</i> ₄	<i>s.e.</i>	<i>b</i> ₅	<i>s.e.</i>
IST	-0.19	0.13	7.54	4.99	-9.35	6.14	x	9.64				
CS1	0.66	0.12	-2.72	0.50	-0.46	0.19	3.23	0.60				
CS2	0.51	0.12	-4.39	1.02	-0.41	0.24	4.42	1.04				
OST	0.40	0.13	-7.85	2.55	2.41	0.81						
GS1	0.22	0.12	-3.09	1.69	6.81	3.65						
GS2	-0.09	0.11	0.05	1.22	x	x	x	x	x	x	x	x
GS3	0.39	0.12	-3.33	1.02	3.08	0.96						
GS4	0.63	0.12	-3.10	0.58	-0.15	0.18	1.50					
GS5	0.87	0.13	-4.03	0.62	-1.90	0.29	0.76					
GS6	2.04	0.24	-2.25	0.21	-1.09	0.11	0.02					
GS7	4.89	1.20	-1.59	0.12	-0.59	0.07	0.34					
GS8	2.04	0.23	-2.07	0.19	-0.91	0.10	-0.10					

There were 2 high (>10) LD values indicating local dependency.

Response patterns of two items (GS1, GS3) had a significant ($p < .05$) lack of fit to their respective model.

Extrinsic time pressure												
Item	<i>a</i>	<i>s.e.</i>	<i>b</i> ₁	<i>s.e.</i>	<i>b</i> ₂	<i>s.e.</i>	<i>b</i> ₃	<i>s.e.</i>	<i>b</i> ₄	<i>s.e.</i>	<i>b</i> ₅	<i>s.e.</i>
IEPT	0.05	0.25	x	x	x	x	x	x				
CEPT	0.72	0.15	-3.59	0.71	-0.92	0.24	1.78	0.36				
OEPT	0.87	0.16	-2.91	0.50	-0.39	0.15	3.15	0.54				
GEP1	1.32	0.25	-3.12	0.48	-1.74	0.25	-0.28	0.11				
GEP2	2.26	0.57	-1.66	0.20	0.10	0.09	1.05	0.14				

Table 9. ... continued

Aversive/Noxious exposures												
Item	<i>a</i>	<i>s.e.</i>	<i>b</i> ₁	<i>s.e.</i>	<i>b</i> ₂	<i>s.e.</i>	<i>b</i> ₃	<i>s.e.</i>	<i>b</i> ₄	<i>s.e.</i>	<i>b</i> ₅	<i>s.e.</i>
INOX1	0.65	0.12	1.17	0.28	2.66	0.52	4.41	0.84				
INOX2	1.37	0.13	-0.44	0.09	0.67	0.11	1.36	0.16				
ONOX1	0.49	0.10	-0.55	0.23	1.09	0.34	2.76	0.63	4.04	0.88		
ONOX2	1.30	0.14	0.56	0.11	1.41	0.17	2.33	0.26				
GNOX1	x	x	-0.62	0.10	-0.12	0.12	0.26	0.01	0.36	0.14		
GNOX2	x	x	-0.62	0.10	-0.12	0.12	0.26	0.01	0.36	0.14		
GNOX3	1.41	0.13	-0.44	0.09	0.79	0.12	1.48	0.16				

There were 7 high (>10) LD values indicating local dependency.

Response patterns of four items (INOX1, INOX2, GNOX1, GNOX2) had a significant ($p < .05$) lack of fit to their respective model.

Symbolic aversiveness												
Item	<i>a</i>	<i>s.e.</i>	<i>b</i> ₁	<i>s.e.</i>	<i>b</i> ₂	<i>s.e.</i>	<i>b</i> ₃	<i>s.e.</i>	<i>b</i> ₄	<i>s.e.</i>	<i>b</i> ₅	<i>s.e.</i>
IAVOI1	x	x	-0.11	x								
IAVOI2	1.82	0.20	-0.41	0.08	0.33	0.09	1.91	0.19				
IAVOI3	0.82	0.13	-1.37	0.22	0.10	0.15	2.40	0.39				
CAVOIT	0.88	0.13	-4.38	0.64	-0.73	0.15	1.32	0.24				
OAVOIT	x	x	-0.12	x	-0.01	x						
GAVOI1	1.10	0.16	0.54	0.14	1.83	0.27						
GAVOI2	1.66	0.20	0.11	0.08	0.78	0.12	1.45	0.17				
GAVOI3	1.20	0.20	1.47	0.23	1.73	0.26	1.99	0.30				
GAVOI4	1.42	0.20	0.63	0.13								
GAVOI5	-0.03	0.12	x	x	-x	x						

There were many high (>10) LD values indicating local dependency.

Response patterns of seven items (IAVOI1, IAVOI3, OAVOIT1, GAVOI2, GAVOI3, GAVOI4, GAVOI5) had a significant ($p < .05$) lack of fit to their respective model.

Conflict												
Item	<i>a</i>	<i>s.e.</i>	<i>b</i> ₁	<i>s.e.</i>	<i>b</i> ₂	<i>s.e.</i>	<i>b</i> ₃	<i>s.e.</i>	<i>b</i> ₄	<i>s.e.</i>	<i>b</i> ₅	<i>s.e.</i>
ICNFL1	x	x	0.13	x	1.15	x						
ICNFL2	x	x	0.13	x	1.15	x						
CCNFL1	1.83	0.19	-2.25	0.18	-0.18	0.07	1.81	0.18				
CCNFL2	1.60	0.17	-1.32	0.13	0.39	0.09	2.46	0.27				
CCNFL3	1.20	0.15	-3.64	0.44	-0.55	0.11	2.16	0.29				
OCNFL1	1.07	0.14	-2.17	0.26	0.04	0.11	2.89	0.41				
OCNFL2	0.62	0.12	-3.85	0.72	0.32	0.21	4.74	1.01				
OCNFL3	0.64	0.12	-1.76	0.33	-0.11	0.17	2.66	0.54				
GCNFL1	0.90	0.13	-2.66	0.36	0.27	0.14	2.63	0.42				
GCNFL2	1.40	0.16	-1.11	0.13	0.59	0.11	2.21	0.26				
GCNFL3	0.29	0.15	3.89	2.04	11.54	5.88						
GCNFL4	1.39	0.16	-1.55	0.16	0.46	0.10	2.49	0.30				
GCNFL5	0.98	0.14	-0.56	0.13	2.28	0.35						
GCNFL6	0.56	0.19	3.40	1.16	3.44	1.18						

There was 1 high (>10) LD value indicating local dependency

Response patterns of four items (ICNFL1, ICNFL2, CCNFL1, OCNFL23) had a significant ($p < .05$) lack of fit to their respective model.

Note. x = absolute value larger than 10.